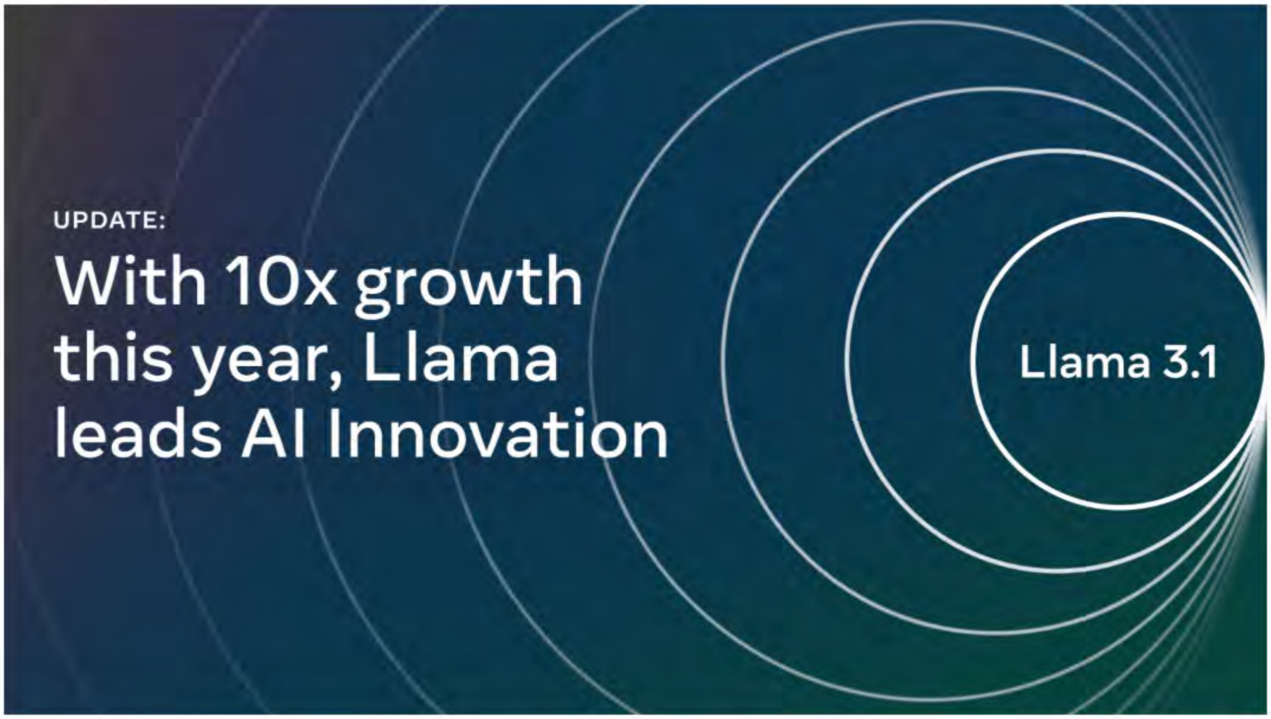
NAYAK
EXHIBIT L-1

Meta Our approach Research Product experiences Llama Blog Try Meta AI

Large language model

With 10x growth since 2023, Llama is the leading engine of AI innovation

August 29, 2024



Key Takeaways:

- Llama models are approaching 350 million downloads to date (more than 10x the downloads compared to this time last year), and they were downloaded more than 20 million times in the last month alone, making Llama the leading open source model family.
- Llama usage by token volume across our major cloud service provider partners has more than doubled in just three months from May through July 2024 when we released Llama 3.1.
- Monthly usage (token volume) of Llama grew 10x from January to July 2024 for some of our largest cloud service providers.

It's been just over a month since we released Llama 3.1, expanding context length to 128K, adding support across eight languages, and introducing the first frontier-level open source AI model with our Llama 3.1 405B. As we did with our Llama 3 and Llama 2 releases, today we're sharing an update on the momentum and adoption we're seeing across the board.

The success of Llama is made possible through the power of open source. By making our Llama models openly available we've seen a vibrant and diverse AI ecosystem come to life where developers have more choice and capability than ever before. The innovation has been broad and rapid, from start-ups pushing new boundaries to enterprises of all sizes using Llama to build on-premises or through a cloud service provider. Industry is building and innovating with Llama, and we're even more excited for what's to come.

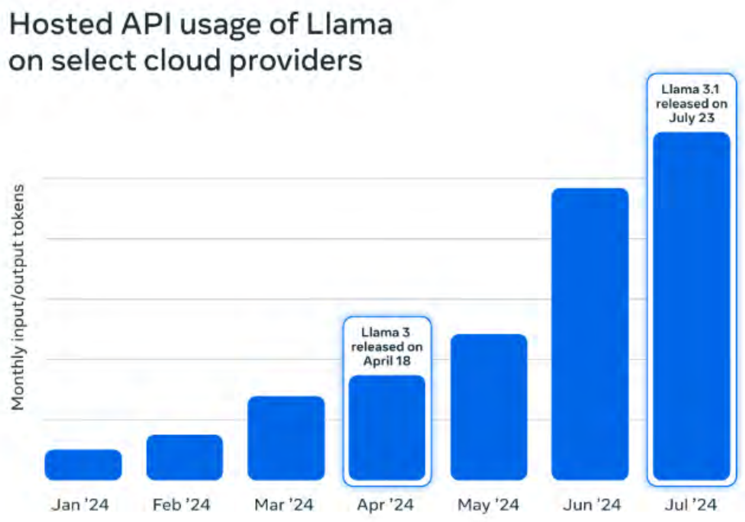
Alongside the release of Llama 3.1, Mark Zuckerberg shared an open letter on the benefits of open source AI—further cementing our vision and commitment to an open approach. Open source is in our company's DNA, and Llama both embodies and reinforces our commitment to sharing our work in a responsible way. Open source promotes a more competitive ecosystem that's good for consumers, good for companies (including Meta), and ultimately good for the world.

In just 18 months since our initial launch, Llama has evolved from a single state-of-the-art foundation model to a robust system for developers. With Llama 3.1, we now offer developers a complete reference system to more easily create their own custom agents along with a new set of security and safety tools to help build responsibly.

The leading open source model

The Llama ecosystem is growing rapidly. Llama models are approaching 350 million downloads on Hugging Face to date—an over 10x increase from where we were about a year ago. Llama models were downloaded more than 20 million times on Hugging Face in the last month alone. And this is just one piece of the Llama success story with these models also being downloaded on services from our partners across the industry.

In addition to Amazon Web Services (AWS) and Microsoft's Azure, we've partnered with Databricks, Dell, Google Cloud, Groq, NVIDIA, IBM watsonx, Scale AI, Snowflake, and others to better help developers unlock the full potential of our models. Hosted Llama usage by token volume across our major cloud service provider partners more than doubled May through July 2024 when we released Llama 3.1.



Monthly usage of Llama grew 10x from January to July 2024 for some of our largest cloud service providers. And in the month of August, the highest number of unique users of Llama 3.1 on one of our major cloud service provider partners was the 405B variant, which shows that our largest foundation model is gaining traction.

We've grown the number of partners in our Llama early access program by 5x with Llama 3.1 and will do more to meet the surging demand from partners. We've heard from a number of companies that want to be future LEAP and integration Llama partners, including Wipro, Cerebras, and Lambda.

Swami Sivasubramanian, VP, AI and Data, AWS: "Customers want access to the latest state-of-the-art models for building AI applications in the cloud, which is why we were the first to offer Llama 2 as a managed API and have continued to work closely with Meta as they released new models. We've been excited to see the uptake for Llama 3.1 from customers across both Amazon SageMaker and Amazon Bedrock, and we look forward to seeing how customers use this model to solve their most complex use cases."

Ali Ghodsi, CEO & Co-Founder, Databricks: "In the weeks since launch, thousands of Databricks customers have adopted Llama 3.1, making it our fastest adopted and best selling open source model ever. This generation of Llama models finally bridges the gap between OSS and commercial models on quality. Llama 3.1 is a breakthrough for customers wanting to build high quality AI applications, while retaining full control, customizability, and portability over their base LLM."

Jonathan Ross, Founder & CEO, Groq: "Open-source wins. Meta is building the foundation of an open ecosystem that rivals the top closed models and at Groq we put them directly into the hands of the developers—a shared value that's been fundamental at Groq since our beginning. To date Groq has provided over 400,000 developers with 5 billion free tokens daily, using the Llama suite of models and our LPU Inference. It's a very exciting time and we're proud to be a part of that momentum. We can't add capacity fast enough for Llama. If we 10x'd the deployed capacity it would be consumed in under 36 hours."

Jensen Huang, Founder & CEO of NVIDIA: "Llama has profoundly impacted the advancement of state-of-the-art AI. The floodgates are now open for every enterprise and industry to build and deploy custom Llama supermodels using NVIDIA AI Foundry, which offers the broadest support for Llama 3.1 models across training, optimization, and inference. It's incredible to witness the rapid pace of adoption in just the past month."

What's even more encouraging than how many people are using Llama is who is using Llama and how they're using Llama.

We're seeing growing preference in the developer community for Llama and strong indicators for continued growth. According to a survey from Artificial Analysis, an independent site for AI benchmarking, Llama was the number two most considered model and the industry leader in open source.

With more than 60,000 derivative models on Hugging Face, there's a vibrant community of developers fine-tuning Llama for their own use cases. Large enterprises like AT&T, DoorDash, Goldman Sachs, Niantic, Nomura, Shopify, Spotify, and Zoom are just a few success stories, and both Infosys and KPMG are using Llama internally.

Let's take a closer look.

A snapshot of Llama case studies

Accenture is using Llama 3.1 to build a custom LLM for ESG reporting that they expect to improve productivity by 70% and quality by 20 – 30%, compared with the company's existing way of generating Accenture's annual ESG report. With its exciting advancements in multilingual capabilities, Accenture is able to extend AI models across regions, for example to help a global organization make chatbots more culturally conscious and relevant. Accenture believes companies will need to leverage many different AI models from different providers. Open source models like Llama 3.1 expand options, accelerate innovation, and will have a positive ripple effect across business and society.

Customer care is an area of focus for AI-powered innovation at **AT&T**. Through fine-tuning Llama models, they've been able to cost effectively improve customer care by better understanding key trends, needs and opportunities to enhance the experience moving forward. Overall, Llama and GenAI have driven a nearly 33% improvement in search-related responses for AT&T customer care engagements while reducing costs and speeding up response times.

DoorDash uses Llama to streamline and accelerate daily tasks for its software engineers, such as leveraging its internal knowledge base to answer complex questions for the team and delivering actionable pull request reviews to improve its codebase.

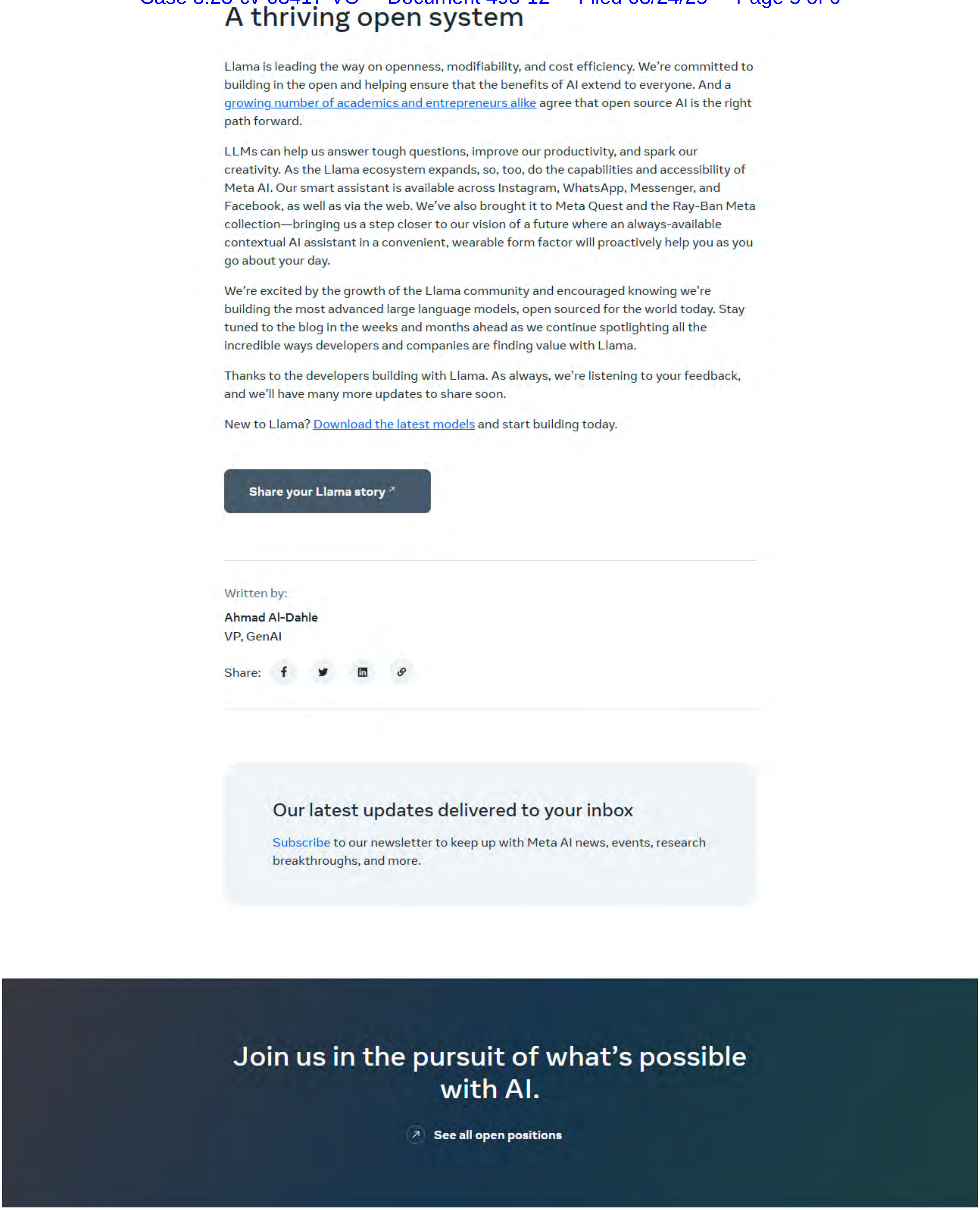
Goldman Sachs AI platform, known as the GS AI Platform, allows Goldman engineers to use Llama models for various use cases in a safe and responsible way, including information extraction from documents.

To drive the virtual world of its first-of-its-kind AR game Peridot, **Niantic** integrated Llama, transforming its adorable creatures, called "Dots," into responsive AR pets that now exhibit smart behaviors to simulate the unpredictable nature of physical animals. Llama generates each Dot's reaction in real time, making every interaction dynamic and unique.

Leading Japanese financial institution **Nomura** uses Llama on AWS for key benefits, including faster innovation, transparency, bias guardrails, and robust performance across text summarization, code generation, log analysis, and document processing.

Shopify is continuing to experiment with best-in-class open source models, including LLaVA, which is built on the foundations of Llama. They use finetunes of LLaVA for multiple specialized tasks and are currently doing 40M – 60M Llava inferences per day supporting the company's work on product metadata and enrichment.

Zoom uses its own models as well as closed- and open-source LLMs—including Llama—to power its AI Companion, a generative AI assistant that helps workers avoid repetitive, mundane tasks. AI Companion serves up meeting summaries, smart recordings, and next steps to Zoom users, freeing up more of their time to collaborate, make connections, and get things done.



A thriving open system

Llama is leading the way on openness, modifiability, and cost efficiency. We're committed to building in the open and helping ensure that the benefits of AI extend to everyone. And a growing number of academics and entrepreneurs alike agree that open source AI is the right path forward.

LLMs can help us answer tough questions, improve our productivity, and spark our creativity. As the Llama ecosystem expands, so, too, do the capabilities and accessibility of Meta AI. Our smart assistant is available across Instagram, WhatsApp, Messenger, and Facebook, as well as via the web. We've also brought it to Meta Quest and the Ray-Ban Meta collection—bringing us a step closer to our vision of a future where an always-available contextual AI assistant in a convenient, wearable form factor will proactively help you as you go about your day.

We're excited by the growth of the Llama community and encouraged knowing we're building the most advanced large language models, open sourced for the world today. Stay tuned to the blog in the weeks and months ahead as we continue spotlighting all the incredible ways developers and companies are finding value with Llama.

Thanks to the developers building with Llama. As always, we're listening to your feedback, and we'll have many more updates to share soon.

New to Llama? Download the latest models and start building today.

Written by:

Ahmad Al-Dahle
VP, GenAI

Share:

Our latest updates delivered to your inbox

Subscribe to our newsletter to keep up with Meta AI news, events, research breakthroughs, and more.

Join us in the pursuit of what's possible with AI.

Related Posts

Computer Vision

Introducing Segment Anything: Working toward the first foundation model for image segmentation

April 5, 2023

→ **Read post**

FEATURED

Meta AI

Research

MultiRay: Optimizing efficiency for large-scale AI models

November 18, 2022

→ **Read post**

FEATURED

ML Applications

MuAViC: The first audio-video speech translation benchmark

March 8, 2023

→ **Read post**

Our approach

About AI at Meta
People
Careers

Research

Infrastructure
Resources
Demos

Search AI content

Product experiences

Meta AI
AI Studio

Latest news

Blog
Newsletter

Foundational models

Llama

Privacy Policy Terms Cookies

Meta © 2025